

DTIC FILE COPY

4

AD-A199 067



Systems  
Optimization  
Laboratory

DTIC  
ELECTE  
SEP 02 1988  
S H D

Department of Operations Research  
Stanford University  
Stanford, CA 94305

DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

4

SYSTEMS OPTIMIZATION LABORATORY  
DEPARTMENT OF OPERATION RESEARCH  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA 94305-4022

PARALLEL PROCESSORS FOR PLANNING  
UNDER UNCERTAINTY

by

George B. Dantzig and Peter W. Glynn

TECHNICAL REPORT SOL 88-8

June 1988

DTIC  
ELECTE  
SEP 02 1988  
S H D

Research and reproduction of this report were partially supported by the National Science Foundation Grants DMS-8420623, ECS-8617905, and SES-8518662; U.S. Department of Energy Grant DE-FG03-87ER25028; Office of Naval Research Contract N00014-85-K-0343, and Electric Power Research Institute Contract RP2940-1; the Center for Economic Policy Research at Stanford University.

Any opinions, findings, an conclusions or recommendations expressed in this publication are those of the author and do NOT necessarily reflect the views of the above sponsors.

Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

88 8 2 068

# PARALLEL PROCESSORS FOR PLANNING UNDER UNCERTAINTY

George B. Dantzig & Peter W. Glynn

## ABSTRACT

In this paper we describe joint research under way by Mordecai Avriel, Robert Entriken, and the authors. Our goal is to demonstrate, for an important class of multistage stochastic models, that a variety of techniques for solving large-scale linear programs can be effectively mixed to attack this fundamental problem. The ideas involve nested primal and dual decomposition, combined with Monte Carlo simulation, high speed importance sampling, and quadrature methods for numerical integration, together with the use parallel processors.

### Keywords:

Linear Programming, Mathematical Programming, Large-Scale Optimization, Deterministic Models, Times-Staged Systems, Staircase Systems, Decomposition Principle, Benders Decomposition, Cutting Planes, Parallel Processors, Stochastic Systems, Reliable Systems, Hedging, Monte Carlo Simulation, Importance Sampling.



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

## 1. Hedging Against Uncertainty

A long outstanding problem of great practical importance concerns finding an efficient way to do planning, scheduling, and control of complex systems under uncertainty. Although progress has been made on this fundamental problem of operations research, control theory, and economics by Roger Wets, John Birge, and others, it remains in general unsolved.

We first state the deterministic version of the problem and then generalize it to include two important characteristics of stochastic problems encountered in practice which we will refer to as *intra-period* and *inter-period* uncertainty. Mathematical programs are used for planning, scheduling, and optimal design of large-scale complex systems. Applications include models used for strategic planning, policy decisions to guide the growth of the economy, scheduling production and expansion of large-scale industrial enterprises such as those that generate and distribute electricity, water, fuel, or produce agricultural products. Many such models have thousands of variables and equations. These models are mostly deterministic. Unfortunately, the solutions of deterministic models are often not taken seriously *because they do not properly hedge against future contingencies*.

While it is relatively easy to reformulate deterministic models to take account of uncertainties, the rub has been that for complex time-staged systems the model size increases exponentially with the number of stages. This has made them too expensive to solve. [10,13]

A variety of heuristic devices are used in practice to adjust deterministic solutions so that they hedge. *Scenario analysis* is one popular way to do this. Several different scenarios, are computed (usually only five or six), the results are compared, and a compromise solution somehow or other is arrived at empirically.

Birge has developed clever ways to arrive at approximate solutions to stochastic programs and ways to estimate the quality of his approximations. [4,5,6].

Our approach is more direct. Many scenarios are run and used to arrive at a compromise solution that hedges against uncertainties. The sample space of all possible scenarios, could be continuous, or could run into millions of discrete points. For many problems, it is reasonable to consider solving thousands of sample scenarios which are used as input data for generating the hedging solution. Since these sample scenarios are independently drawn, it is easy to see why parallel processors are ideal for efficiently carrying out such computations.

One class of deterministic models, which we generalize to the stochastic case are the time-staged linear programming models whose matrix structure is lower block triangular, [11,12,14]. Other basic references are [2,10,11,15,33]. See also [18,26,28,29,31,32]. By introduction of in-process inventories and other devices, this class can be reduced to the mathematically equivalent "staircase" problems of the form:

FIND min  $Z$  and vectors  $X_t > 0$ , such that

$$\begin{aligned}
 b_1 &= A_1 X_1 \\
 b_2 &= -B_1 X_1 + A_2 X_2 \\
 &\vdots \\
 b_t &= -B_{t-1} X_{t-1} + A_t X_t \\
 &\vdots \\
 b_T &= -B_{T-1} X_{T-1} + A_T X_T \\
 (\min) Z &= c_1 X_1 + \dots + c_t X_t + \dots + c_T X_T
 \end{aligned}$$

where matrices  $A_t, B_t$  and vectors  $b_t, c_t$  are given.

Suppose one of several contingencies (events) can happen in the second period so that  $b_2, B_1$ , and  $A_2$  are not known with certainty. We index the possible events by  $\omega = 1, 2, \dots, K$  and assume that  $p(\omega)$ , the probability of the event  $\omega$  is known. Then in place of the second relation

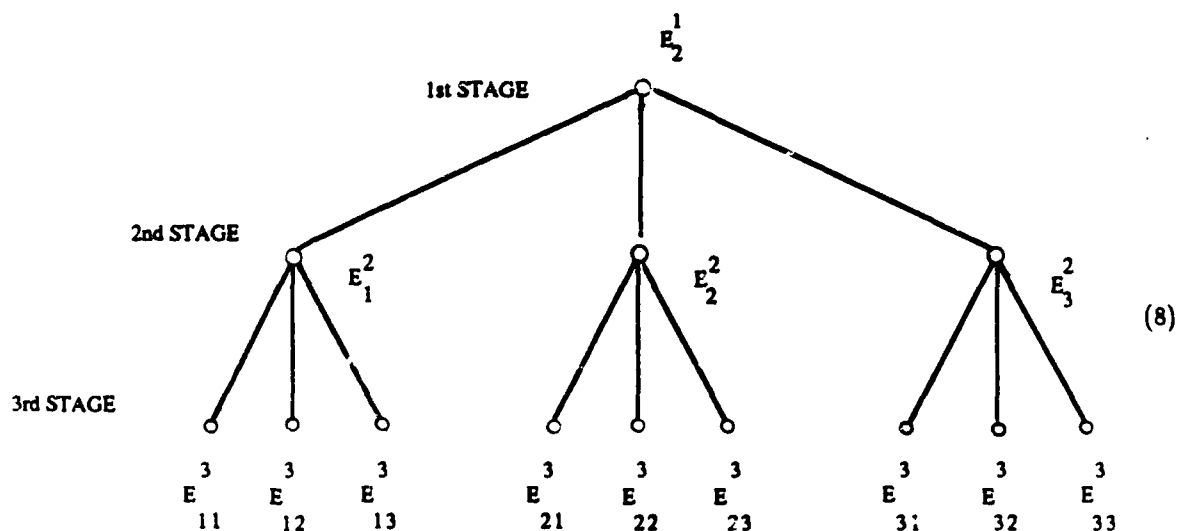
$$b_2 = -B_1 X_1 + A_2 X_2$$

we have many relations of the form

$$\begin{aligned}
 b_2 &= -B_1(1)X_1 + A_2(1)X_2(1) \\
 &\vdots \\
 b_2(\omega) &= -B_1(\omega)X_1 + A_2(\omega)X_2(\omega) \\
 &\vdots \\
 b_2(K) &= -B_1(K)X_1 + A_2(K)X_2(K)
 \end{aligned}$$

If there are no further contingencies after the second period, then associated with each  $X_2(\omega)$  will be a system of relations associated with it of exactly the same form as those below  $X_2$  above except variables  $X_t$  for  $t \geq 2$  are replaced by  $X_t(\omega)$ . In the objective equation, the terms  $c_t X_t$  are replaced by their expected values,  $c_t \sum p(\omega) X_t(\omega)$ .

In general, however, there will be contingencies happening in every period. The "event tree" of contingent events in this more general case has the form



The number of branches associated with each node can be finite or infinite. It is obvious why the size of the system can grow exponentially with the number of stages. Moreover, even if a problem has only two stages, there can be a large or infinite number of possible contingencies in the second stage. Some researchers, like Roger Wets, [33,34,35] have concentrated on the two stage case because it is an important problem in its own right and because it can be used as a stepping stone for finding solutions to the multi-stage case for certain classes of problems as we will soon see.

## 2. The Intra-Period Stochastic Submodel

The general stochastic programming appears to us to be intractable given the present state of the art. We, therefore, have been concentrating on classes of models that are relevant and whose event tree does not grow exponentially with the number of time periods being modelled, [13]. Here are some examples:

Typically, industries use their facilities to carry out operations. Thus an airline has a fleet of airplanes of different kinds, and has other facilities for handling passengers on the ground and repairing aircraft. Their operations consist of flying aircraft, maintaining them, and serving passengers on the ground. In the case of an electric utility, it has facilities for generating and distributing electricity (dams, generators using nuclear fuel, fossil fuel, water power, and transmission lines). Operationally, these facilities are used to generate and distribute electricity. [16]

Planning models for such industries may be essentially deterministic as far as their plans for expansion of facilities are concerned. In the airline example, the deterministic part are the schedules for purchase or retirement of aircraft in the fleet and the expansion of ground facilities. Their operations, however, must be modeled in a stochastic way in order to be sure the facilities on hand are sufficient to take care of various contingencies that might arise in day-to-day operations.

What allows us to decompose the problem into a deterministic part and uncertainty part is the assumption that, whatever be the contingencies that arise in day to day operations, the facilities will not be destroyed in the process of using them for operations. (A situation in which this might appear not be true would be an aircraft being used for operations which is destroyed by an accident. This contingency, however, can be modeled so as not to affect the future state of facilities if there is insurance to cover the loss.) Models of this type for two periods have the form:

Let  $\omega = 1, 2, \dots$  and  $\bar{\omega} = 1, 2, \dots$  be independently drawn random variables.

FIND  $\min Z$ ,  $X_t \geq 0$ ,  $U_1(\omega) \geq 0$ ,  $U_2(\bar{\omega}) \geq 0$  such that:

$$\begin{aligned} b_1 &= A_1 X_1 \\ *** \quad d_1(\omega) &= -D_1(\omega) X_1 + F_1(\omega) U_1(\omega) \\ b_2 &= -B_1 X_1 + A_2 X_2 \\ *** \quad d_2(\bar{\omega}) &= -D_2(\bar{\omega}) X_2 + F_2(\bar{\omega}) U_2(\bar{\omega}) \\ \min Z &= c_1 X_1 + c_2 X_2 \\ &\quad + \sum_{\omega} p_1(\omega) g_1(\omega) U_1(\omega) + \sum_{\bar{\omega}} p_2(\bar{\omega}) g_2(\bar{\omega}) U_2(\bar{\omega}) \end{aligned}$$

The probabilities that  $\omega$  and  $\bar{\omega}$  occur are given by  $p_2(\omega), p_2(\bar{\omega})$ . The objective minimizes expected costs. The *Intra-period Submodels* correspond to the set of equations marked by \*\*\* above, one for each period. The remaining equations constitute the "Deterministic" part of the system.

Benders decomposition is an ideal way to solve such a model as we shall soon see, [3,17]. Under this approach, the submodels are solved with  $X_1$  and  $X_2$  temporarily fixed at some values,  $X_1 = X_1^*$  and  $X_2 = X_2^*$ . Note that, when  $X_1$  and  $X_2$  are known, the submodels decompose into many, many small independent subproblems, one for each value of  $\omega = 1, 2, \dots$ , and  $\bar{\omega} = 1, 2, \dots$ , namely

$$\text{FIND } \min g_1(\omega) U_1(\omega), U_1(\omega) \geq 0, \text{ such that } F_1(\omega) U_1(\omega) = d_1(\omega) + D_1(\omega) X_1^* ;$$

and for each  $\bar{\omega} = 1, 2, \dots$ , namely

$$\text{FIND } \min g_2(\bar{\omega}) U_2(\bar{\omega}), U_2(\bar{\omega}) \geq 0, \text{ such that } F_2(\bar{\omega}) U_2(\bar{\omega}) = d_2(\bar{\omega}) + D_2(\bar{\omega}) X_2^* .$$

Parallel processors can be effectively used to solve these problems wholesale for all choices of  $\omega$  and  $\bar{\omega}$  when there are not too many values of  $\omega$  and  $\bar{\omega}$ . When there are too many  $\omega$  and  $\bar{\omega}$ , then a "representative" sample is used instead.

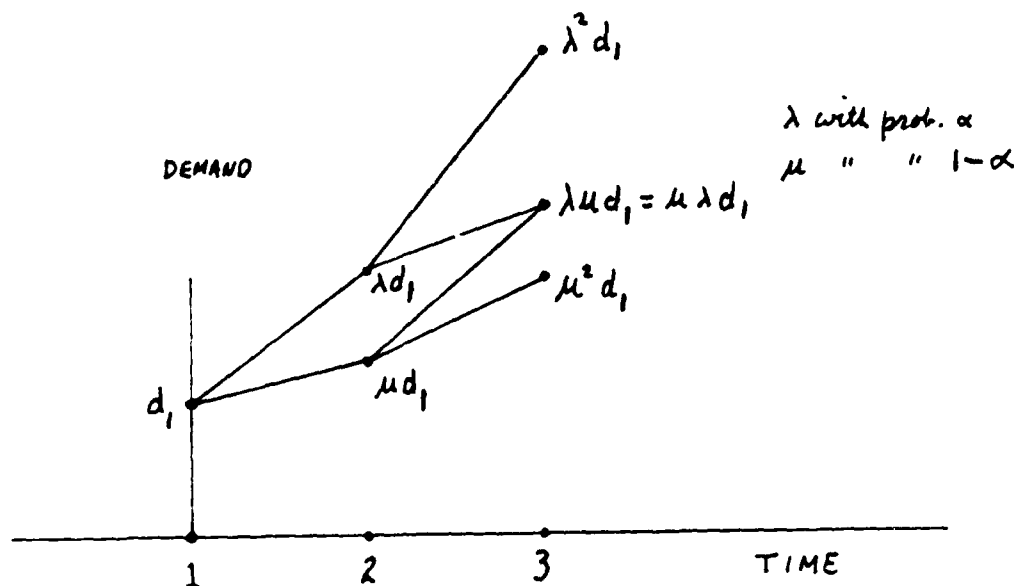
### 3. The Inter-Period Stochastic Part of the model (Ideas in this section are due to Moredecai Avriel).

We now consider an equally relevant class of models where the contingencies that arise in one period affect later periods. Suppose in one year the demand for some item is high and this higher

demand it more likely that demand in the following year will also be higher. The simplest case would be one in which demand  $d_{t+1}$  in period  $t + 1$  is related to the demand in period  $t$  by

$$d_{t+1} = \begin{cases} \lambda d_t & \text{with probability } \alpha \\ \mu d_t & \text{with probability } 1 - \alpha = \beta \end{cases}$$

where  $\lambda > \mu$ .



In this simple example, the number of cases at time  $t$  will be  $2^t$ , i.e., the size of the problem is growing exponentially in  $t$ . However, because  $\lambda \mu d_1 = \mu \lambda d_1$ , there is a consolidation in the number of cases so that in fact, the number of cases is only growing proportional to  $t$ .

Even though a low demand in year  $t - 2$  followed by a high demand in year  $t - 1$  might arrive with the same probability of an intermediate demand in period  $t$  as a high demand in period  $t - 2$  followed by a low demand in period  $t - 1$ , this does not mean we can equate other components in the state space. For models having a more complex structure, the state at time  $t$  of the other components of the system could be quite different depending on the history of their past states. Mordecai Avriel of our research team has proposed a way to reduce the number of cases in general by forcing the consolidation of the case HL generated by a high ( $= H$ ) followed by a low ( $= L$ ) year and the case LH generated by a low ( $= L$ ) followed by high ( $= H$ ) year. He recommends averaging the other components of their state vector at time  $t$ . This is illustrated below for a three period model:



FIND  $\min Z, (X_1, X_2^L, X_2^H, X_3^{LL}, X_3^{LH}, X_3^{HH}) \geq 0$ :

$$\begin{array}{llll}
 AX_1 & & & = b_1 \\
 *** & & & *** \\
 -B^L X_1 + A_2^L X_2^L & & & = B_2^L \\
 *** & & & *** \\
 -B_1^H X_1 & + A_2^H X_2^H & & = b_2^H \\
 & *** & & *** \\
 -B_2^{LL} X_2^L & + A_3^{LL} X_3^{LL} & & = b_3^{LL} \\
 & *** & & *** \\
 -\frac{1}{2} B_2^{LH} X_2^L - \frac{1}{2} B_2^{HL} X_2^H & + A_3^{LH} X_3^{LH} & & = b_3^{LH} \\
 & *** & & *** \\
 -B_2^{HH} X_2^H & + A_3^{HH} X_3^{HH} & & = b_3^{HH} \\
 & *** & & *** \\
 \hline
 c_1 X_1 & + \alpha c_2^L X_2^L & + (\beta) c_2^H X_2^H & + \alpha^2 c_3^{LL} X_3^{LL} + 2\alpha\beta c_3^{LH} X_3^{LH} + \beta^2 c_3^{HH} X_3^{HH} = Z(\min)
 \end{array}$$

The asterisks \*\*\* above indicate that the intra-period stochastic constraints have been omitted. The first and second set of omitted relations are:

$$\begin{aligned}
 d_1(\omega) &= -D_1(\omega)X_1 + F_1(\omega)U_1(\omega), \quad \omega = 1, 2, \dots \\
 d_2^L(\bar{\omega}) &= -D_2^L(\bar{\omega})X_2^L + F_2^L(\bar{\omega})U_2^L(\bar{\omega}), \quad \bar{\omega} = 1, 2, \dots
 \end{aligned}$$

#### 4. Solving the Inter-period part of the Stochastic Model

Without the intra-period \*\*\* constraints, when the number of time periods is small, it may be practical to use standard linear programming software to directly solve the model. A model having too many \*\*\* constraints may nevertheless be tractable if we replace operating constraints by "cuts" generated when Benders decomposition is used to find tentative solutions to the subproblems, [3]. These cuts, for example, are many inequalities of the form

$$\begin{aligned}
 \varepsilon \pi_1(\omega) d_1(\omega) &\leq -\varepsilon \pi_1(\omega) D_1(\omega) X_1 + \theta_1, \quad \omega = 1, 2, \dots \\
 \varepsilon \pi_2^H(\bar{\omega}) d_2(\bar{\omega}) &\leq -\varepsilon \pi_2^H(\bar{\omega}) D_2^H(\bar{\omega}) X_2^H + \theta_2^H, \quad \bar{\omega} = 1, 2, \dots
 \end{aligned}$$

When the size of the inter-period part of the model becomes too large to solve directly using standard linear programming software, it is planned to use *nested* Benders decomposition software. Such software has already been developed by Robert Entriken for solving staircase systems by modifying the MINOS linear programming code, [19,23,24,25].

To apply Benders decomposition, the model is partitioned by columns — see dotted vertical lines [1]. The first "Master" corresponds to the columns in the first period. The Master assigns

a tentative value to  $X_1$ , say  $X_1 = X_1^*$ . The subproblem corresponds to the remaining columns and is solved assuming that  $X_1 = X_1^*$  is specified and terms  $-B_1^L X_1^*$  and  $B_2^H X_1$  have been added to the right hand side. This subproblem can, in turn, be partitioned into a Master and Sub with the Master corresponding to variables  $X_2^L, X_2^H$ . (The earlier consolidation of two states in period 3 complicates the discussion which follows.) To simplify matters assume, instead, either the term  $B_2^{LH} X_2^L$  or  $B_2^{HL} X_2^H$  is used to represent the LH or HL states but not both. If this is done the Benders subproblem at each stage will turn out to be a set of smaller problems to be iteratively solved, of which the following is typical;

$$\begin{aligned} A_2^L X_2^L &= b_2^L + B_1 X_1^*, & X_2^L &\geq 0 \\ \text{CUTS} \quad -G_2^\ell X_2^L + \theta &\geq g_2^\ell, & \ell &= 1, 2, \dots \\ c_2 X_2^L + \theta &= \min \end{aligned}$$

The theory of how such cuts are generated and how information is passed back and forth to other subproblems will be outlined next.

#### 4.0. Solving the Intra-Period Stochastic Model

We begin with the simplest *two-stage case* first studied in Dantzig [10] and developed by R. Wets [33,34,35]:

$$\begin{aligned} b_1 &= A_1 X_1, & (X_1, X_2) &\geq 0, \\ b_2 &= -B_1 X_1 + A_2 X_2 \\ (\min) Z &= c_1 X_1 + c_2 X_2 \end{aligned} \tag{3}$$

where the first stage  $(b_1, A_1, c_1)$  are known with certainty while the second stage  $(b_2, c_2, B_1, A_2)$  is assumed to be functions of a random variable  $\omega$  with known probability distribution  $p(\omega), \omega \in \Omega$ . The values of  $\omega$  in  $\Omega$  may have a continuum of values in which case  $p_2(\omega)$  is a probability density distribution; or  $\omega$  may take on a finite or an infinite set of discrete values, in which case  $p_2(\omega)$  is a discrete probability distribution where  $\omega = 1, 2, \dots, K$  where  $K$  may be infinite. When applied to the intra-period submodel discussed earlier, the equation  $b_2 = -B_1 X_1 + A_2 X_2$  is replaced by  $d_1(\omega) = -D_1(\omega) X_1 + F_1(\omega) U_1(\omega)$  with corresponding changes in the objective form;

Let  $\Omega$  be the sample space of  $\omega$ . For the purposes of the computational approach outlined in this paper, we require that  $\Omega$  be discrete with a finite number of elements. Practically speaking, this is no restriction since any distribution may be approximated by a probability mass function concentrated on a finite set of points. Then, assuming we label the sample points  $\omega$  using the integers  $\{1, 2, \dots, k\}$ , the random vectors and matrices  $(b_2, c_2, B_1, A_2)$  takes on the value  $(b_2(\omega), c_2(\omega), B_1(\omega), A_2(\omega))$ ,  $(1 \leq \omega \leq K)$  with known probability  $p_2(\omega)$ .

We now illustrate the approach for  $K = 2$ . The stochastic problem of minimizing expected costs under uncertainty then has *as its certainty equivalent* the deterministic linear program that we outlined earlier except now we describe the computational method in greater detail:

$$\text{Find min } Z, \quad X_1 \geq 0, \quad X_2(\omega) \geq 0, \quad \omega = 1, 2, 3 : \quad (4.0)$$

$$b_1 = A_1 X_1 \quad (4.1)$$

$$b_2(1) = -B_1(1)X_1 + A_2(1)X_2(1) \quad (4.2)$$

$$b_2(2) = -B_1(2)X_1 + A_2(2)X_2(2)$$

$$b_2(3) = -B_1(3)X_1 + A_2(3)X_2(3)$$

$$\text{min } Z = c_1 X_1 + p_2(1) c_2(1) X_2(1) + p_2(2) c_2(2) X_2(2) + p_2(3) c_2(3) X_2(3) \quad (4.3)$$

To simplify the discussion, assume a bounded optimal solution exists. It follows that we can always find  $\pi_2(\omega)$  to premultiply constraints corresponding to  $b_2(\omega)$  above and subtract from the objective so that adjusted  $c_2(\omega) \geq 0$ . Therefore we can assume without loss of generality  $c_2(\omega) \geq 0$ . Except as noted otherwise, we will assume  $B_1$  is independent of  $\omega$ , i.e.,  $B_1 = B_1(\omega)$  for all  $\omega$ .

Typically, as we have already noted, this problem is solved using "Benders" decomposition, see [3]. The key idea is to replace the contribution of the second period variables to the objective function by a scalar  $\theta_2$ , and to replace the second period constraints — those shown in (4.2) between the dashed lines — by a set of inequalities expressed in terms of  $X_1$  and  $\theta_2$  only, called "cuts". These are *necessary conditions* which are satisfied by all feasible and optimal solutions to (4). These cuts, are added sequentially ( $\ell = 1, 2, \dots$ ) to the first period problem,  $A_1 X_1 = b_1, X_1 \geq 0$ . And these, together with a modified objective  $Z = c_1 X_1 + \theta$  constitute the "Restricted MASTER Problem" whose  $Z$  is a lower bound estimate for  $\min Z$  of (4.1) ... (4.3). Cuts are added to the Master until they become *sufficient* to solve (4). This happens when the current value of the objective  $Z$  for a feasible solution to (4) equals the lower bound estimate of  $\min Z$ . In practice the iterative process is stopped when this difference is judged to be "small enough". Cuts come in two "flavors": *feasibility cuts* and *optimality cuts*. The "MASTER" problem for Benders' decomposition method has the form:

$$\text{FIND } \min Z, \quad X_1 \geq 0, \quad \theta_2 \geq 0 :$$

$$b_1 = A_1 X_1, \quad (5.1)$$

$$\text{CUTS: } g_1^\ell \leq -G_1^\ell X_1 + \delta_2^\ell \theta_2, \quad \ell = 1, \dots, L \quad (5.2)$$

$$\min Z = c_1 X_1 + \theta_2 \quad (5.3)$$

where  $\delta_2^\ell = 0$  for feasibility cuts if the subproblem from which it was derived is infeasible, and  $\delta_2^\ell = 1$  for optimality cuts if the subproblem (6) below is feasible. The optimal solution  $X_1 = X_1^*$  to (5.0) ... (5.3) is the value of  $X_1$  that is temporarily specified and passed to the subproblem where it is "tested" to see if it qualifies as the first period component of some optimal solution  $[X_1, X_2(1), X_2(2), X_2(3)]$  for (4). This is done by solving the set of subproblems (6) below to see (i) if the contribution  $B_1 X_1^*$  from the first period implies for the second period a feasible solution

for every choice of  $\omega$ , and (ii) if it together with the set of optimal solutions to the second period for every  $\omega$  provides a global optimum to the original problem. Global optimality is easily tested by checking whether the lower bound estimate for  $\min Z$  is equal to the value of  $Z$  for the current feasible solution. If the answer to (i) or (ii) is negative, the optimal  $\pi_2(\omega)$  to (6) is substituted in formula (7.1) or (7.2) below in order to generate cut  $L + 1$  which is then added to the  $L$  already generated in (5.2). It can be proved that the current optimal solution of the Master violates the cut condition and therefore the next optimal solution to the Master will generate an improved lower for  $Z$ , [3].

#### 4.1 The Sub Sub Problem

For each  $\omega$  in  $\Omega$ , FIND  $\min Z_2(\omega), X_2(\omega) \geq 0$ :

$$\begin{array}{ll} A_2(\omega)X_2(\omega) & = b_2(\omega) + B_1X_1^* \\ p_2(\omega) \cdot c_2(\omega)X_2(\omega) & = Z_2(\omega)(\min) \end{array} \quad \begin{array}{l} \text{Dual Prices} \\ : \pi_2(\omega) \end{array} \quad (6)$$

where  $\omega = \{1, \dots, K\}$ . These problems are solved for  $\omega = 1, \dots, K$  and their optimal dual "prices" (if (6) is feasible), or "infeasibility" prices (if not feasible) are computed and used as follows: If any subproblem  $\omega$  is infeasible, its infeasibility prices are used to generate a "feasibility" cut (7.1) below with  $\delta_2^{\ell+1} = 0$

$$g_1^{\ell+1} = \pi_2(\omega)b_2(\omega); \quad G_1^{\ell+1} = \pi_2(\omega)B_1(\omega). \quad (7.1)$$

If feasible for all  $\omega \in \Omega$ , then  $X_1^*$  is tested for optimality by comparing the lower bound estimate of  $\theta$  from the master problem with  $\sum_{\omega} Z_2(\omega)$ . If the test fails the expected values:

$$g_1^{\ell+1} = \sum_{\omega} \pi_2(\omega)b_2(\omega); \quad G_1^{\ell+1} = \sum_{\omega} \pi_2(\omega)B_1(\omega). \quad (7.2)$$

are used to generate new "optimality" cut conditions to augment those of (5.2) with  $\delta_2^{\ell+1} = 1$ . Note that (7.2) are actually *expected* values because  $\pi_2(\omega)$  as defined by (6) is proportional to the probabilities  $p_2(\omega)$ .

#### 5. The Concept of Reliable Systems

The stochastic operations submodels above have been formulated so that facilities made available for day-to-day operations are always sufficient to meet the demands on the system whatever be the contingency  $\omega$ . Formulating the model this way can make the facilities required too costly to build. Instead of requiring the system to be *always feasible*, it is often formulated to be *reliable*, i.e., feasible most of the time.

Conditions that place an upper bound on the allowed frequency of failure to meet demand turn out to be non-convex when expressed in terms of the usual variables representing the levels

of operations and therefore cannot be approximated satisfactorily in a linear programming context, [16]. However, conditions that measure the *expected amount of demand not satisfied* are easy to express linearly, one such constraint being added to each stage. The submodels for each state for  $\omega = 1, 2, \dots$  will no longer be independent. The way to restore independence is to make these extra constraints correspond to a "Super Master" (in the Dantzig-Wolfe primal decomposition sense). The Super Master systematically assigns penalty weights to the extra conditions and these are used to modify the objective. If this approach is adopted, independence of the subproblems is restored; the subproblems are reformulated so that they are always feasible whatever be  $\omega$ , [11].

## 6. Using Parallel Processors

The decomposition algorithm, however, is clearly only practical when  $K$  is small. When  $K$  is large, it is proposed that parallel processors be used as high-speed sampling or quadrature devices to effectively solve the subproblems. One idea is to have a processor at the MASTER level serve as an *integrator* which sequentially receives as input estimates of the cuts (5.2). The Master Problem is then solved to optimality with the estimates it has received so far and used to generate as output, revised  $X_1 = X_1^*$  that are sent to other parallel processors which are busy solving (6) for various choices of  $\omega$ . This process also provides a *lower bound* estimate for  $\min Z$  which monotonically increases with each solution of the master problem.

The amount of space needed to store the generated cuts in the computer memory need not be high. Assuming  $B_1$  is independent of  $\omega$ , no more than  $L \leq m_2$  of the cuts will be tight on any major iteration, where  $m_2$  is the number of rows in  $B_1$ . This is so because  $G_1^\ell$ , generated by linear combinations of the rows of  $B_1$ , has rank  $\leq r$  where  $r \leq m_2$  is the rank of  $B_1$ . The remainder may be dropped (possibly to be regenerated on some later iteration).

Several parallel processors could be at the SUB level, each having as input the latest value of  $X_1^*$  and solving (6) in dual form for many random or stratified choices of  $\omega$ . When  $c_2, A_2$  are the same for all  $\omega$ , the dual of (6) is a linear program with only the dual objective  $b_2(\omega)$  changing. By judiciously stratifying the random sampling of  $\Omega$  we hope to use the optimal basic dual feasible solution for one  $\omega$  to find quickly the optimal one for the next  $\omega$ . To provide cuts for the MASTER, the parallel processors are to be used to determine the expected values  $g_1^{\ell+1}$  and  $G_1^{\ell+1}$  defined by (7.2) or to approximate them by means of a large enough "importance sample", see Section 7.

If it is practical to solve (6) for all  $\omega$ , the set of solutions to (6) generates a valid cut and a correct *lower bound* estimate for  $\min Z$ . In that case, the difference between the lower bound and upper bound estimates can then be used to test optimality of  $X_1^*$  for the original problem. When  $Z$ , according to some specified tolerance, is close enough to the lower bound estimate for  $\min Z$  the iterative process is stopped and  $X_1^*$  declared "optimal".

## 7. Importance Sampling

For the cases where  $K$  is large, it is no longer possible to solve (6) for all  $\omega$ . Instead, we propose to use random sampling to choose a set of  $\omega$ 's for which (6) will be solved. [20,21] This solution strategy will require

- a) the development of an efficient sampling plan,
- b) the development of an efficient stopping rule.

By a), we refer to the fact that *naive sampling* as a computational tool, will tend to be inefficient in the sense that a large number of  $\omega$ 's will typically be needed to obtain a reasonable degree of solution accuracy. The reason why this is so for the class of applications we have in mind is that certain  $\omega$ 's play a particularly important role in the solution. For example, in an electric utility capacity planning problem, the  $\omega$ 's corresponding to generator or transmission line failure, while comprising only a small portion of the total sample space  $\omega$ , are significant enough contingencies to force the utility to "hedge". Hence, it is important to design sampling schemes which concentrate an appropriate level of computational effort on these "rare"  $\omega$ 's. We will use two basic ideas, from Monte Carlo simulation, to accomplish this task: stratification and importance sampling. [7,22,27].

In *stratification*, one pre-assigns a certain proportion of the total sample to each of (say)  $m$  subsets partitioning the sample space  $\Omega$ . This increases the efficiency of the sampling procedure by reducing the clustering effects typical of a conventional sampling scheme. For example, in naive random sampling, the entire sample could (with small probability, of course) fall into one subset. There is also a variant of stratification which we will be considering, called *pre-stratification*, that is easier to program, see Cochran, W.G. [9].

The second concept that we shall exploit is *importance sampling*. Within each subset of the stratification partition, we can design our sampling procedure so that we sample not according to the original probability mass function (or, more precisely, the original mass function conditioned on  $\omega$  belonging to the particular subset), but rather according to a mass function which assigns more weight to the "important" elements of the sample space. By "important", we mean those elements which will contribute significantly to the average value of the dual variables. The estimator needs to be appropriately adjusted to account for the new sampling mechanism, but this is easily done, see Hammersley and Handscomb [22]. We intend to use both theory and exploratory data analysis to guide us in developing efficient importance sampling-algorithms.

As for problem (6) described above, we will need to develop a stopping rule (hopefully sequential) which meshes appropriately with the mathematical programming ideas described elsewhere. Specifically, the stopping rule should ensure that a sufficient accuracy is obtained at each iteration of the sampling procedure so as to impart useful information to the optimization loop of the routine. We expect the basic structure of the stopping rule to be of Chow-Robbins type, see [8].

The above sampling ideas should prove to have powerful applications in the optimization con-

text of interest here. We believe in this approach to be fundamental for two reasons, one historical and the other prospective. First, the dimension of the sample space over which expectations need to be computed usually is huge. Monte Carlo methods are easy to extend to the parallel computing environment, and the speed-ups are significant, [30]. The reason, of course, is that Monte Carlo methods are based on replication and replication is trivial to distribute over many parallel processors. For both the above reasons, we believe that Monte Carlo ideas, in conjunction with the mathematical programming concepts developed for solving large-scale systems on main frames, form a promising avenue for the development of efficient solution algorithms for complex stochastic optimization problems.

## 8. REFERENCES

- [1] Abrahamson, P.G. (1983). A nested decomposition approach for solving staircase linear programs, Report SOL 83-4, Department of Operations Research, Stanford University, Stanford, California.
- [2] Beale, E.M.L., Dantzig, G.B. and Watson, R.D. (1986). A First Order Approach to a Class of Multi-Time-Period Stochastic Programming Problems, **Mathematical Programming Study** 27, pp. 103-117.
- [3] Benders, J.F. (1962). Partitioning procedures for solving mixed-variable programming problems, **Numerische Mathematik** 4, pp. 238-252.
- [4] Birge, J.R. (1984). Aggregation in Stochastic Linear Programming, **Mathematical Programming** 31, pp. 25-41.
- [5] Birge, J.R. and S.W. Wallace (1988). A Separable Piecewise Linear Upper Bound for Stochastic Linear Programs, **SIAM J. Control and Optimization** 26 No. 3.
- [6] Birge, J.R. and J.B. Wets (1986). Designing Approximation Schemes for Stochastic Optimization Problems, in Particular for Stochastic Programs with Recourse, **Mathematical Programming Study** 27, pp. 54-102.
- [7] Bratley, P., Fox, B., and Schrage, L. (1983). A Guide to Simulation. Springer-Verlag, New York.
- [8] Chow, Y.S. and H. Robbins (1965). On the Asymptotic Theory of Fixed Width Sequential Confidence Intervals for the Mean, **Ann. Math. Stat.** 36, pp. 457-462.
- [9] Cochran, W.G. (1977). **Sampling Techniques**, John Wiley, New York.
- [10] Dantzig, G.B. (1955). Linear programming under uncertainty, **Management Science**, 1, pp. 197-206.
- [11] Dantzig, G.B. (1963), **Linear Programming and Extensions**, Princeton University Press, Princeton.
- [12] Dantzig, G.B. (1982). Time-staged methods in linear programs, in Studies in Management Science and Systems, Vol. 7 **Large Scale Systems** Y.Y. Haimi (ed.). North Holland Publishing Company, Amsterdam 1982, pp. 19-30.
- [13] Dantzig, G.B. (1988). Planning under uncertainty using parallel computing, **Annals of Operations Research**, in press.
- [14] Dantzig, G.B., Dempster, M.A.H. and Kallio, M.J. (eds.) (1981). **Large-Scale Linear Programming (Volumes 1, 2)**, IIASA Collaborative Proceedings Series, CP-81-51, IIASA, Laxenburg, Austria.



- [15] Dantzig, G.B. and M. Madansky (1961). On the Solution of Two-Staged Linear Programs under Uncertainty, **Proceedings Fourth Berkeley Symposium on Mathematical Statistics and Probability I**, J. Neyman (ed.), pp. 165-176.
- [16] Dantzig, G.B. and M.V.F. Pereira, et al. (1988). Mathematical Decomposition Techniques for Power System Expansion Planning, EPRI EL-5299, Volumes 1 - 5, Electric Power Research Institute, Palo Alto, CA.
- [17] Dantzig, G.B. and Wolfe, P. (1960). The decomposition principle for linear programs, **Operations Research** 8, pp. 110-111.
- [18] Emoliev, Y. (1983), Stochastic Quasigradient Methods and Their Applications to Systems Optimization, **Stochastics** 9, pp. 1-36.
- [19] Glassey, R. (1973). Nested decomposition and multi-stage linear programs, **Management Science** 20, pp. 282-292.
- [20] Glynn, P.W. and W. Whitt (1988). Efficiency of Simulation Estimates. Submitted for publication.
- [21] Glynn, P.W. and D.L. Iglehart (1988). Importance Sampling for Stochastic Simulation. Submitted for publication.
- [22] Hammersley, J.M. and D.C. Handscomb (1964). **Monte Carlo Methods**, Mathuen, London.
- [23] Ho, J.K. and Loute, E. (1980). An advanced implementation of the Dantzig-Wolfe decomposition algorithm for linear programming, Discussion Paper 8014, Center for Operations Research and Econometrics (CORE), Belgium.
- [24] Ho, J.K. and Manne, A.S. (1974). Nested decomposition for dynamic models, **Mathematical Programming** 6, pp. 121-140.
- [25] Jackson, P.L. and D.F. Lynch (1982). Revised Dantzig-Wolfe Decomposition for Staircase-Structured Linear Programs, Technical Report 558, School of Operations Research and Industrial Engineering, Cornell University (revised 1985). To appear in **Mathematical Programming**.
- [26] Kall, P. (1979). Computational Methods for Solving Two-Stage Stochastic Linear Programming Problems, **Z. Angew. Math. Phys.** 30, pp. 261-271.
- [27] Lavenberg, S.S. and Welch P.D. (1981). A perspective on the use of control variables to increase the efficiency of Monte Carlo simulation. **Management Science**, 27, pp. 322-335.
- [28] Louveaux, F.V. (1986). Multistage stochastic programs with block-separable recourse, **Mathematical Programming Study** 28, pp. 48-62.
- [29] Nazareth, L. and R.J-B. Wets (1986). Algorithms for Stochastic Programs: The Case of Nonstochastic Tenders, **Mathematical Programming Study** 28, pp. 1-28.

- [30] Niederreiter, H. (1986). Multidimensional numerical integration using pseudo random numbers, **Mathematical Programming Study** 27, pp. 17-38.
- [31] Prekopa, A., (1978). Dynamic Type Stochastic Programming Models, in **Studies in Applied Stochastic Programming**, (A. Prekopa, ed.), Hungarian Academy of Science, Budapest, pp. 179-209.
- [32] Strazicky, B. (1980). Computational Experience with an Algorithm for Discrete Recourse Problems, in **Stochastic Programming**, (M. Dempster, ed.), Academic Press, London, pp. 263-274.
- [33] Wets, R.J. (1966). Programming under uncertainty: the equivalent convex program, **J. SIAM Applied Math.** 14 (1), pp. 89-105.
- [34] Wets, R.J. (1984). Large-scale linear programming techniques in stochastic programming, **Numerical Methods for Stochastic Optimization**, (Y. Ermoliev and R. Wets, eds.), Springer-Verlag. IIASA WP-84-90.
- [35] Wets, R.J. (1985). On Parallel Processor Design for Solving Stochastic Programs, in **Proceedings of the 6th Mathematical Programming Symposium**, Japanese Mathematical Programming Society, Tokyo, pp. 13-36.

# UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report SOL 88-8	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Parallel Processors for Planning Under Uncertainty		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) George B. Dantzig and Peter W. Glynn		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0343
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research - SOL Stanford University Stanford, CA 94305-4022		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 1111MA
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research - Dept. of the Navy 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE June 1988
		13. NUMBER OF PAGES Pages 16
		14. SECURITY CLASS. (of this report) UNCLASSIFIED
		15. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) This document has been approved for public release and sale; its distribution is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Linear Programming, Mathematical Programming, Large-Scale Optimization, Deterministic Models, Times-Staged Systems, Staircase Systems, Decomposition Principle, Benders Decomposition, Cutting Planes, Parallel Processors, Stochastic Systems, Reliable Systems, Hedging, Monte Carlo Simulation, Importance Sampling.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Please see Abstract on reverse side...		

# ABSTRACT

In this paper we describe joint research under way by Mordecai Avriel, Robert Entriken, and the authors. Our goal is to demonstrate, for an important class of multistage stochastic models, that a variety of techniques for solving large-scale linear programs can be effectively mixed to attack this fundamental problem. The ideas involve nested primal and dual decomposition, combined with Monte Carlo simulation, high speed importance sampling, and quadrature methods for numerical integration, together with the use parallel processors.